Lending Club Case Study

Prepared By: Sreelesh PS, G Bhargav

Presentation Outline

- Executive Summary
- Abstract & Problems
- Business Objectives
- Data Source
- Methodology
- Exploratory Data Analysis
- Evaluation/Results

Lending Club

Executive Summary

In summary, this company serves as the largest online loan marketplace, providing a range of financial products including personal loans, business loans, and medical procedure financing. The primary financial challenge it faces is credit loss, primarily driven by borrowers who default on their loans, labeled as 'charged-off' customers. To mitigate this risk, the company has embarked on a case study aimed at identifying high-risk loan applicants through exploratory data analysis (EDA). By understanding the critical variables associated with loan defaults, the company can better assess risk and manage its portfolio effectively, ultimately enhancing its financial sustainability and performance.

Abstract & Problems

Abstract

This case study addresses a critical business problem faced by a consumer finance company specializing in lending various types of loans to urban customers. The primary issue pertains to making loan approval decisions based on applicant profiles, balancing the risks of potential financial loss. The company aims to minimize two types of risks: denying loans to applicants who would repay (resulting in lost business) and approving loans to applicants who might default (leading to financial losses). To tackle this challenge, the study employs Exploratory Data Analysis (EDA) to understand the influence of consumer and loan attributes on the likelihood of loan default.

Business Problem

The company must balance loan approvals to minimize financial loss by identifying high-risk applicants through EDA for effective risk assessment and portfolio management.

Analytics Problem

The main challenge is conducting EDA on a loan dataset to identify key variables indicating loan default potential, offering insights for informed loan decisions.

Business Objectives

Reduce credit loss by identifying and cutting loans to high-risk applicants.

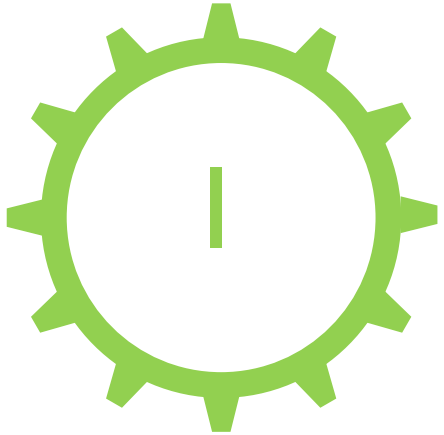
Enhance portfolio management and risk assessment with key driver variables.

Understand factors behind loan defaults via EDA for financial sustainability.

Offer easy, low-interest loans through an online platform as a market leader.

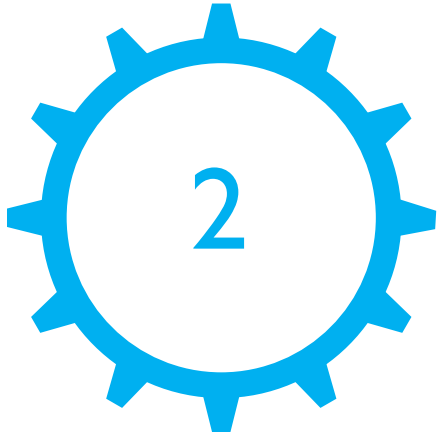
Data Source

The data is downloaded from the Upgrad. It contains the complete loan data for all loans issued through the time period 2007 to 2011.



Loan.csv

It contains the complete loan data for all loans issued through the time period 2007 t0 2011.



Data Dictionary.xlsx

data dictionary which describes the meaning of these variables

Methodology

1. Data Collection

- Gather relevant data sources and datasets for analysis.
- Ensure data integrity, consistency, and quality.

2. Data Cleaning

- Handle missing values, outliers, and data inconsistencies.
- Standardize data formats and remove noise.

3. Data Visualization

- Create charts, graphs, and plots to visualize data distributions and relationships.
- Explore patterns and trends within the data.

4. Descriptive Statistics

- Compute summary statistics (mean, median, standard deviation, etc.) for key variables.
- Identify central tendencies and variability in the data.

5. Data Preprocessing

- Normalize or scale data as necessary.
- Encode categorical variables into numerical format (if applicable).

6. Exploratory Data Analysis

- Investigate relationships and correlations between variables.
- Identify potential outliers and anomalies.

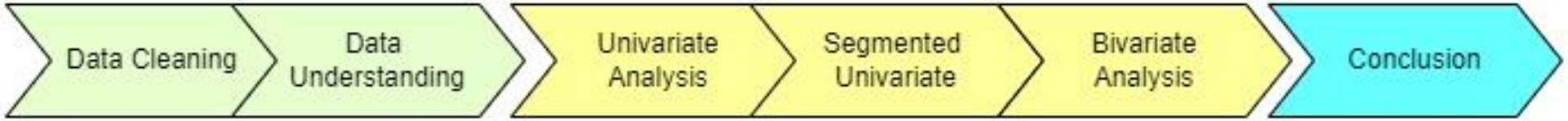
7. Data Interpretation

- Draw conclusions from the EDA findings.
- Make informed decisions or recommendations based on the analysis.

8. Visual Communication

- Prepare clear and insightful visualizations and reports.
- Communicate the EDA results to stakeholders effectively.

EDA Approach



Removing null values unnecessary variables and checking null value percentage

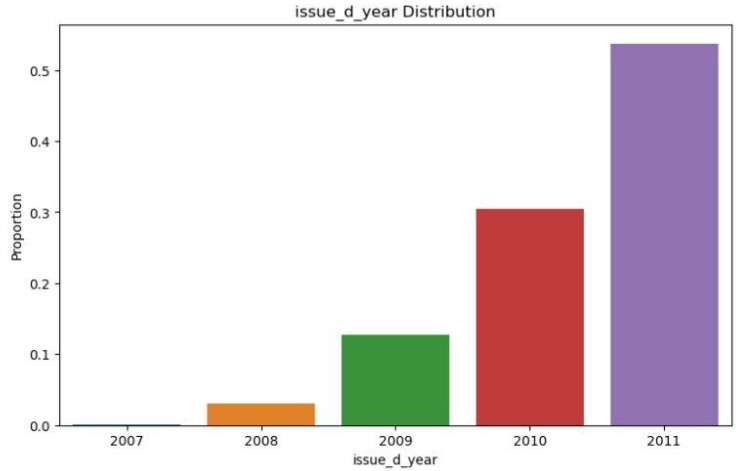
Working with data dict to understand columns

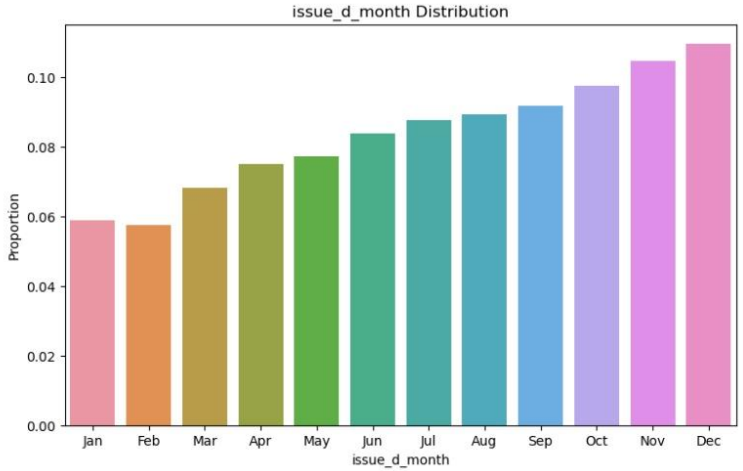
Analysing each column and plotting distribution.
Analysing data columns with categorical data
Analysing multiple variable behaviour and plotting patterns

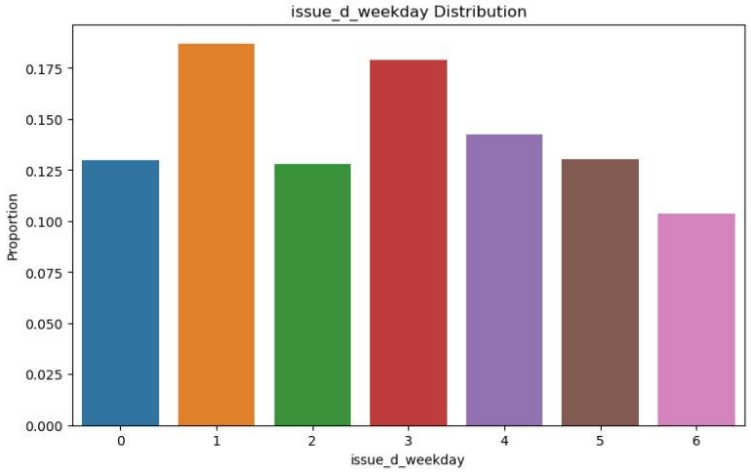
Concluding recommendations for business by analysing all plots.

Exploratory Analysis – Key Findings

Growth of Lending Club



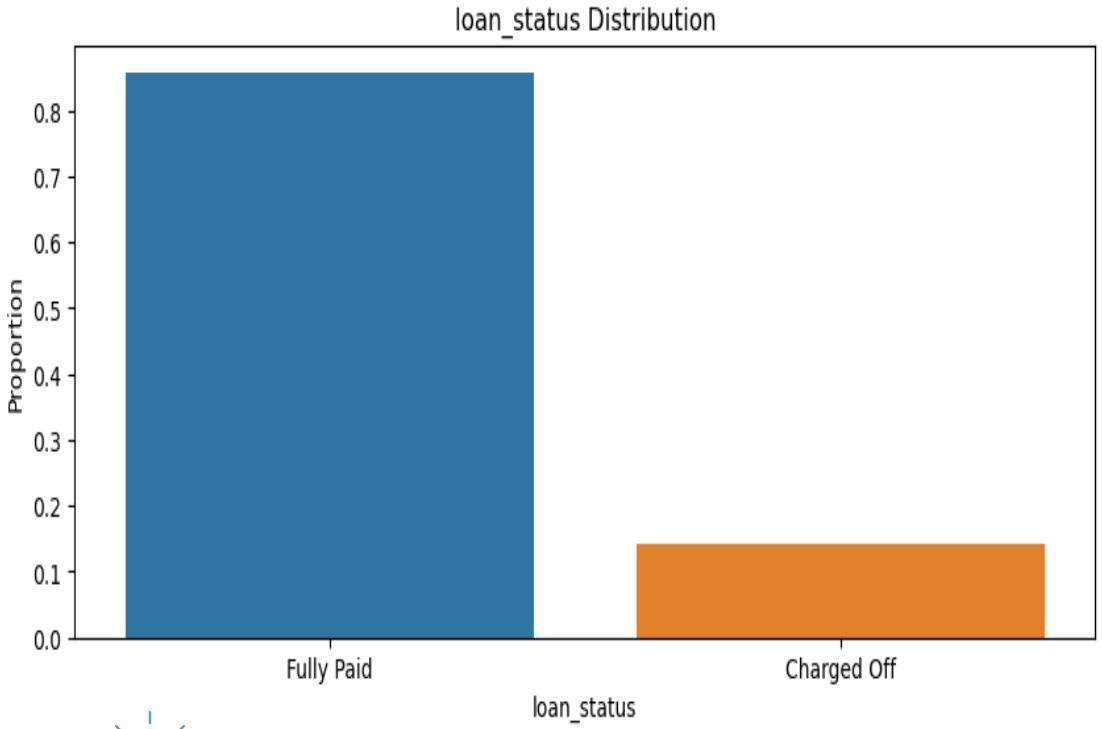




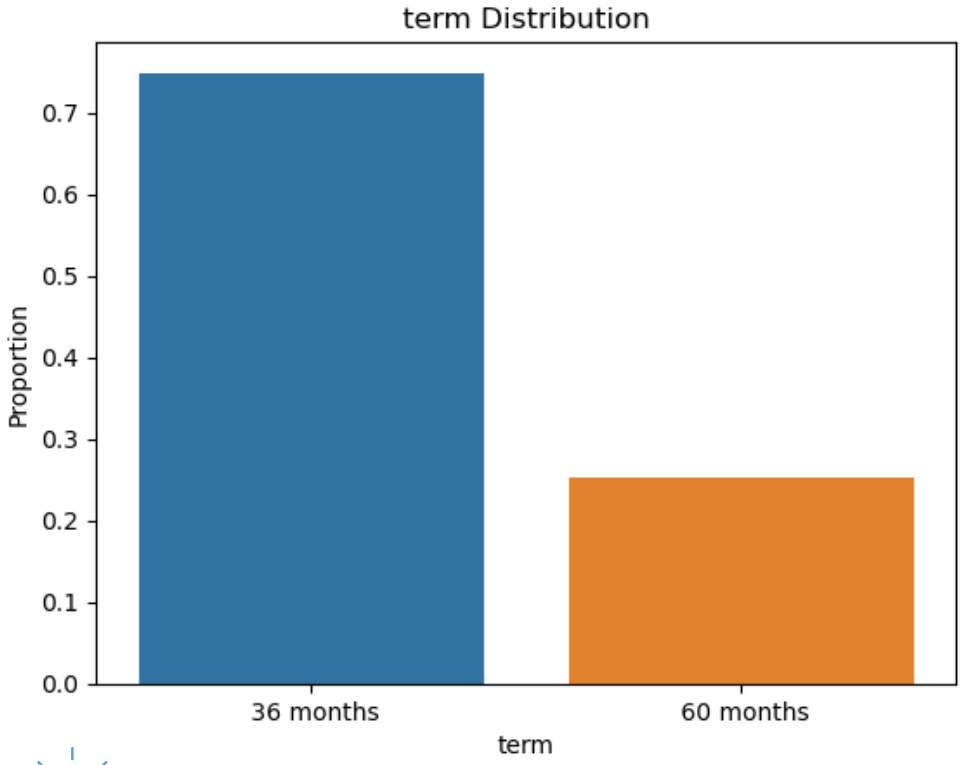
- Lending Club has doubled its expansion in issuing loans every year
- The last two quarters of the year are busier compared to the first two quarters.

Exploratory Analysis – Key Findings

Lending Club Loan & Term Status

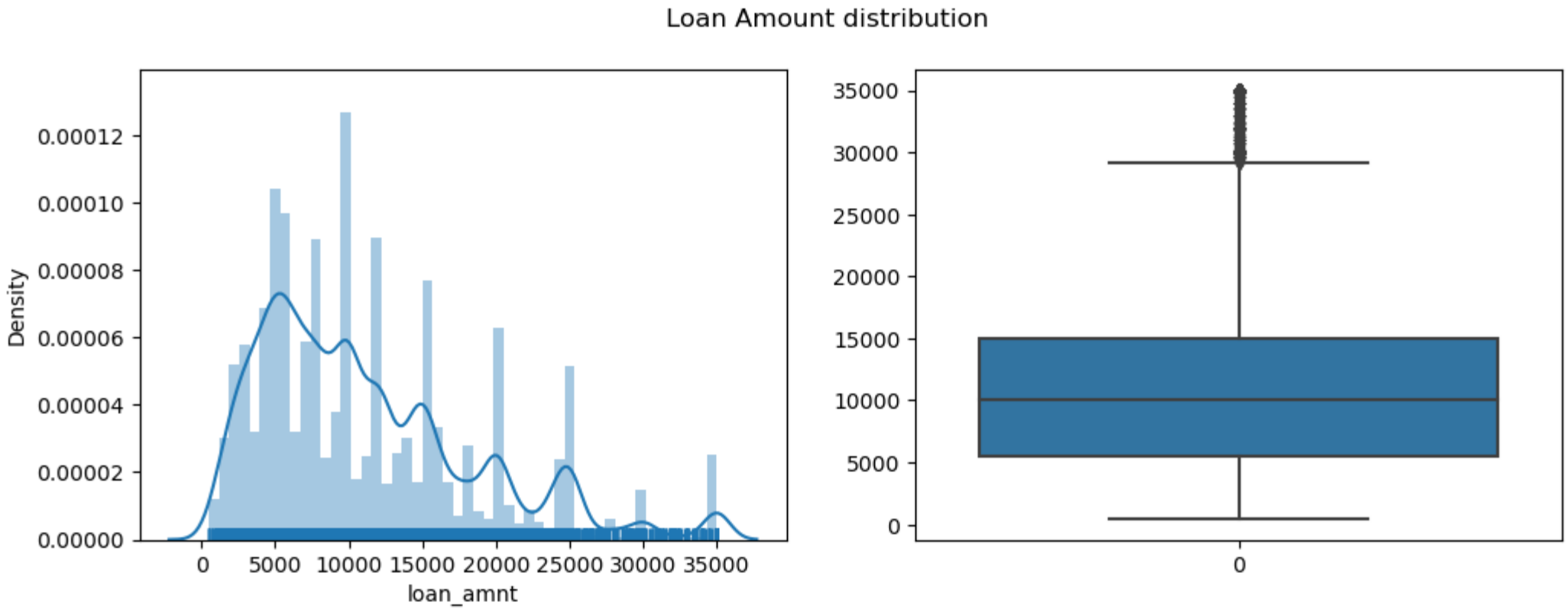


Around 85% of the loans issued are Fully paid
15% of loans issued are defaulted



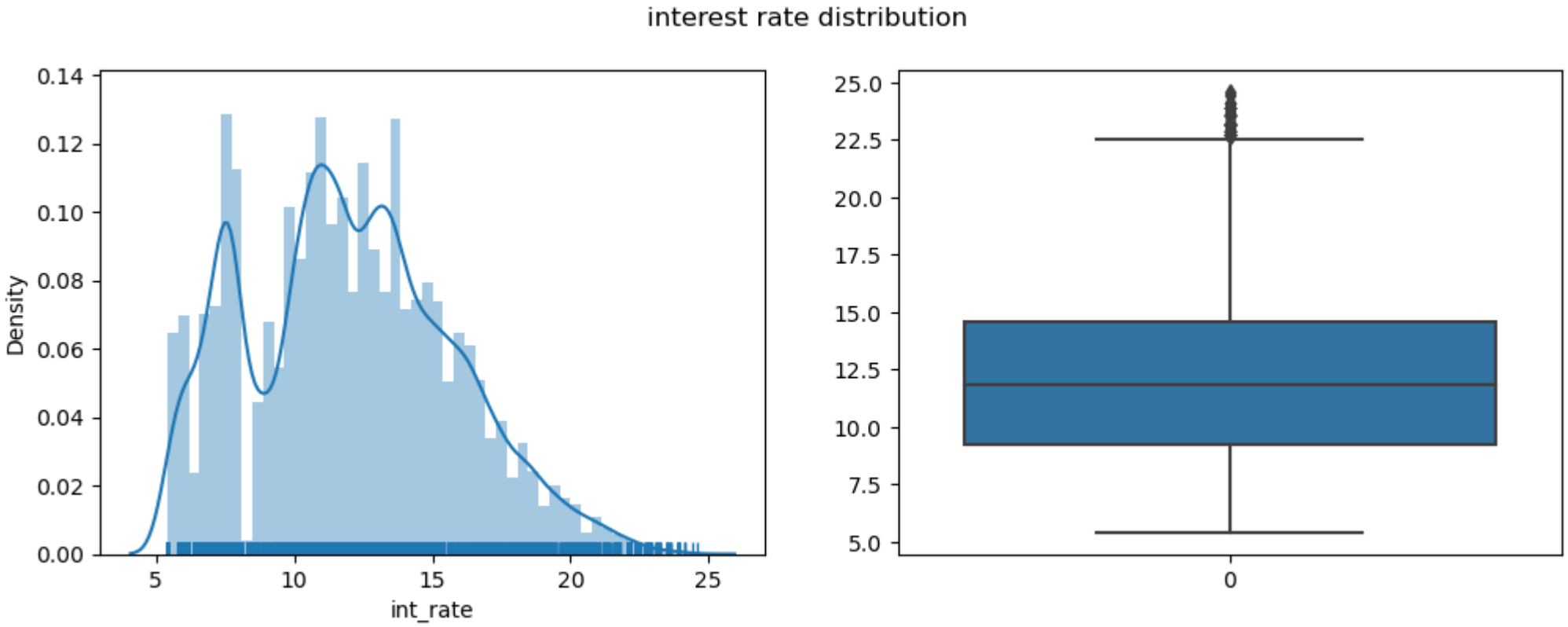
Lending Club issue loans only on 2 terms.
36 months and 60 months

Exploratory Analysis – Key Findings



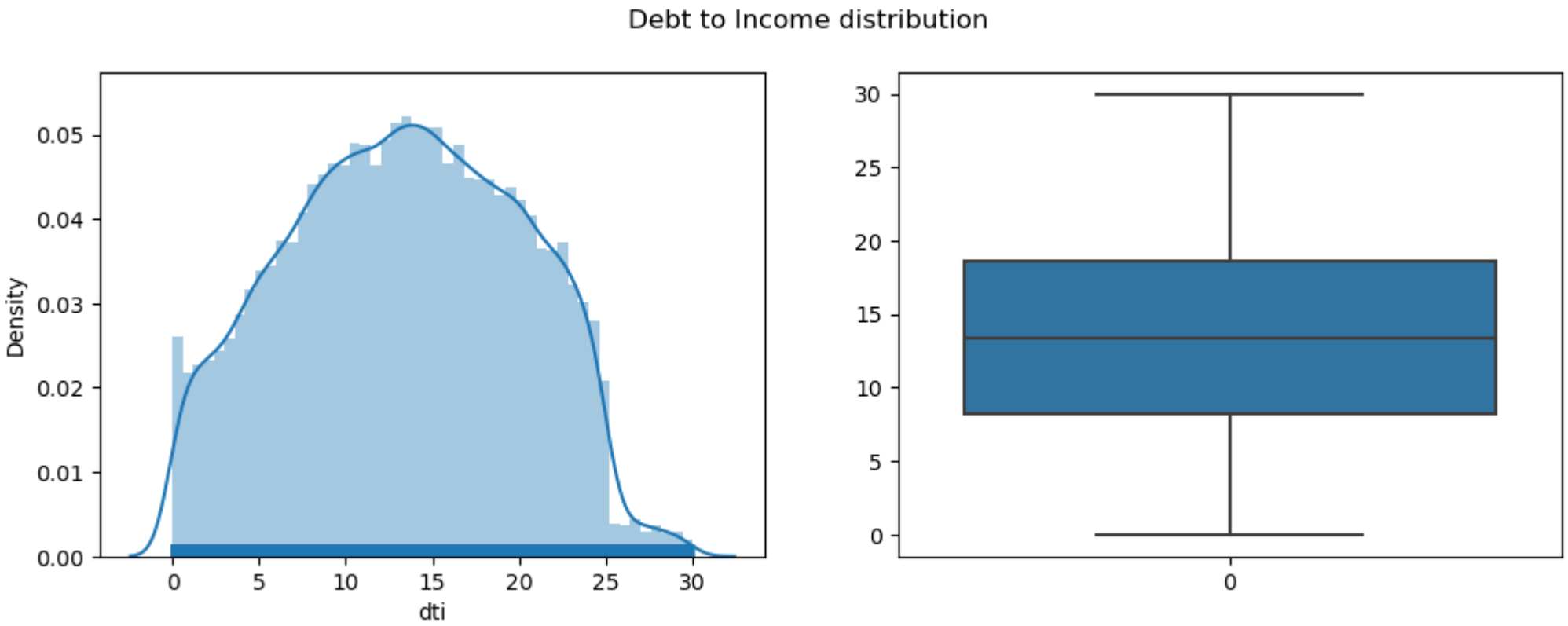
From the above analysis and charts, The Loan amount range from 500 - 35000 , most no.of loans issued are around the amount 10000

Exploratory Analysis – Key Findings



From the above observation we can conclude that most of the intrest rate range between 9-14.5%, where some loans are issued at higher 22.5%, this can be a outlier.

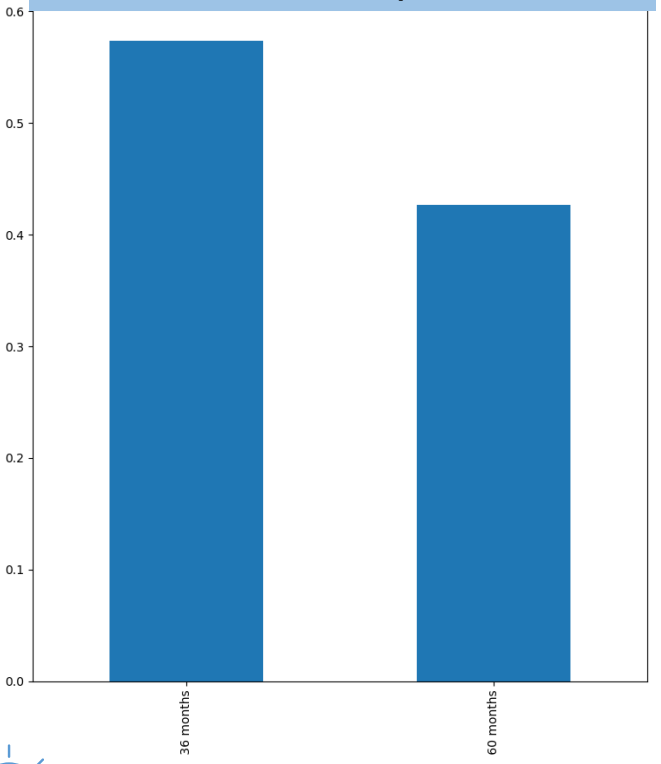
Exploratory Analysis – Key Findings



From the above distribution it looks like a normal distribution and most of the borrower's dti is less than 30

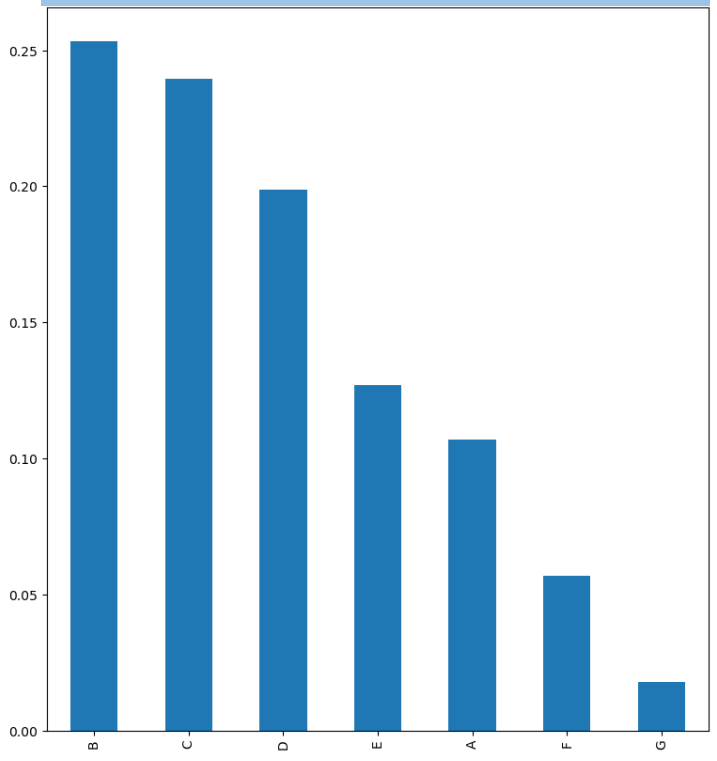
Exploratory Analysis – Key Findings

Defaulters % by Term



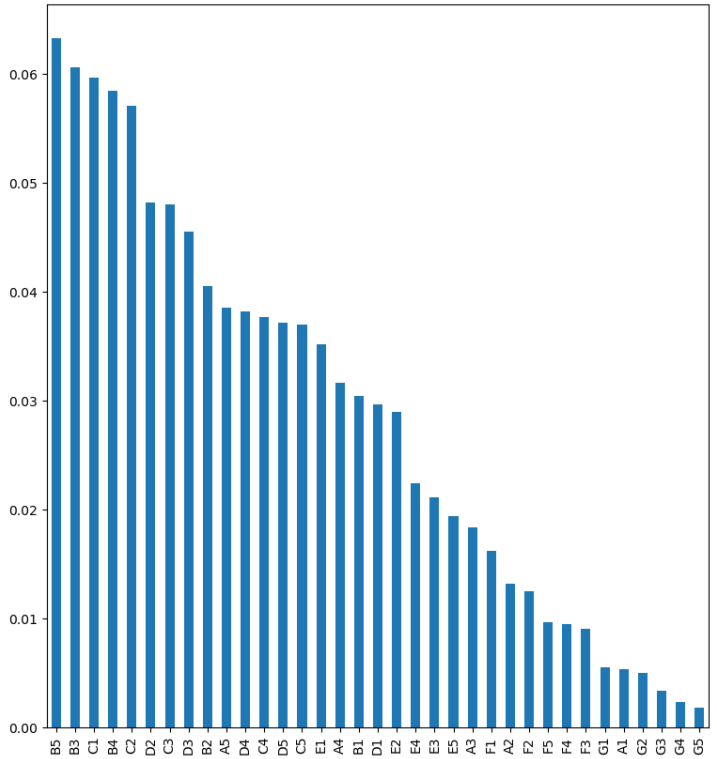
Loans with a 36-month term have a higher default rate compared to those with a 60-month term.

Defaulters % by Grade



Loans with grades B, C, and D have a higher default rate compared to loans with other grades.

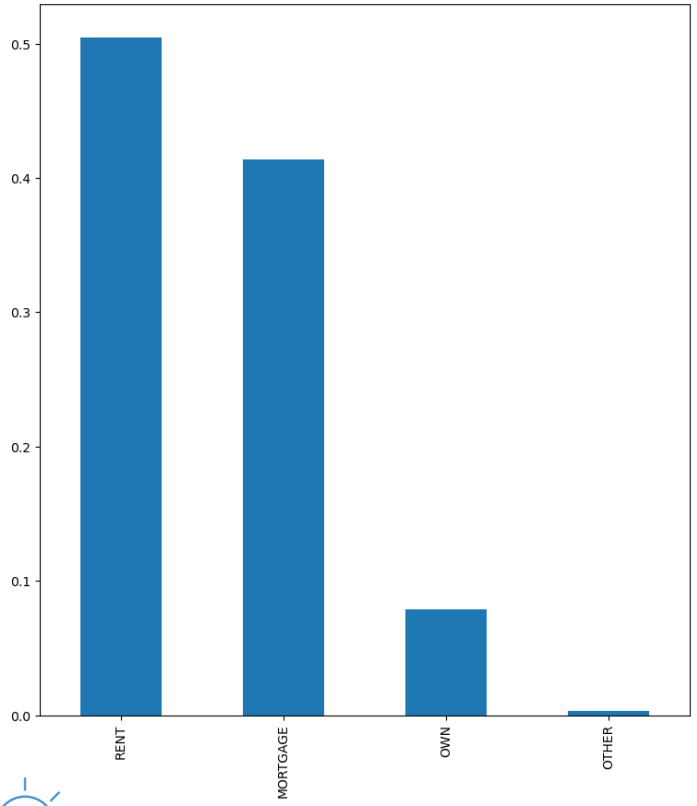
Defaulters % by Sub Grade



The top 5 subgrades with the highest default rates are B5, B3, C1, B4, and C2.

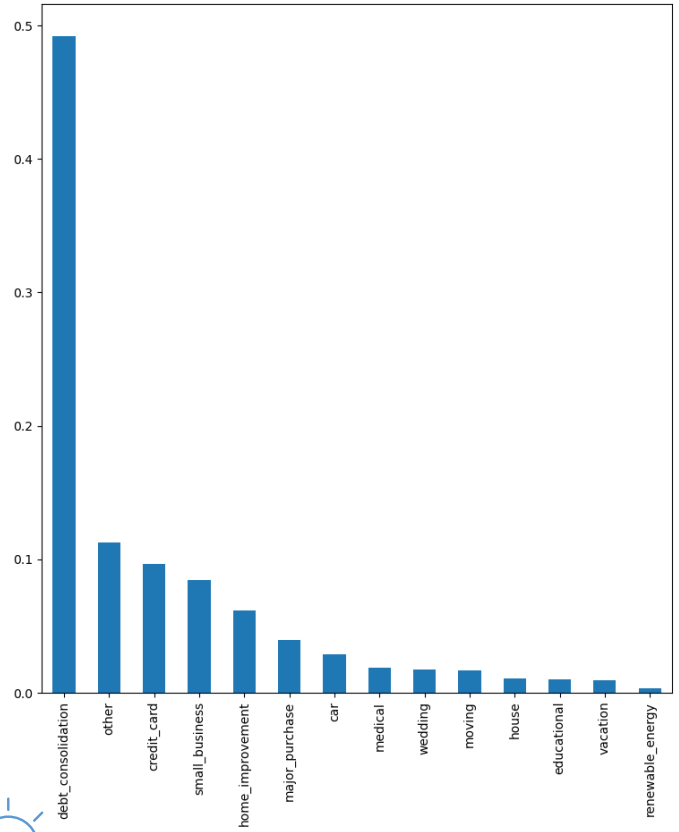
Exploratory Analysis – Key Findings

Defaulters % by Home ownership



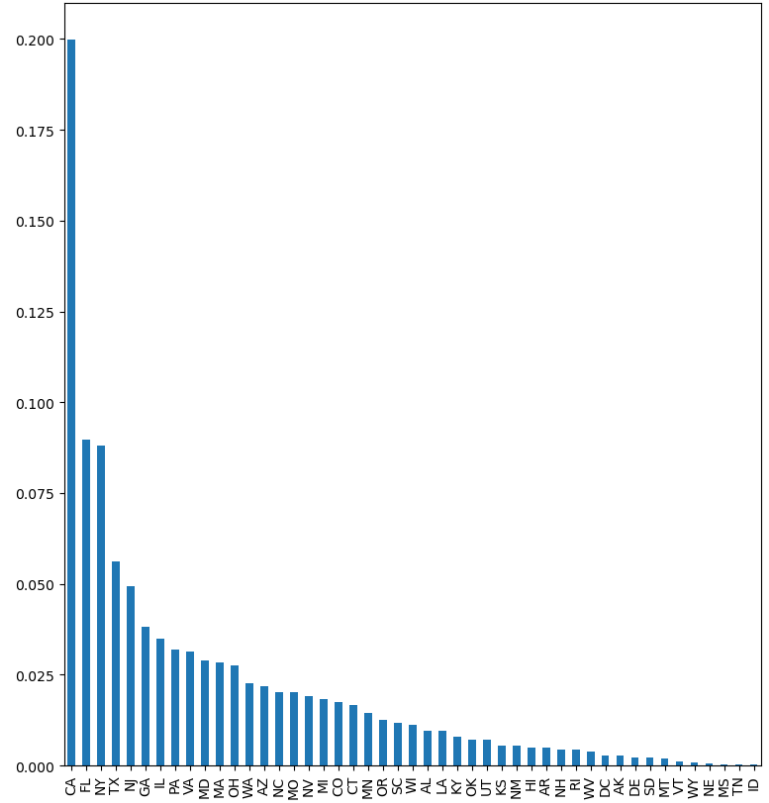
Borrowers who own their homes have a lower default rate compared to those who do not own a home.

Defaulters % by purpose



Loans taken for the purpose of debt consolidation have the highest default rate.

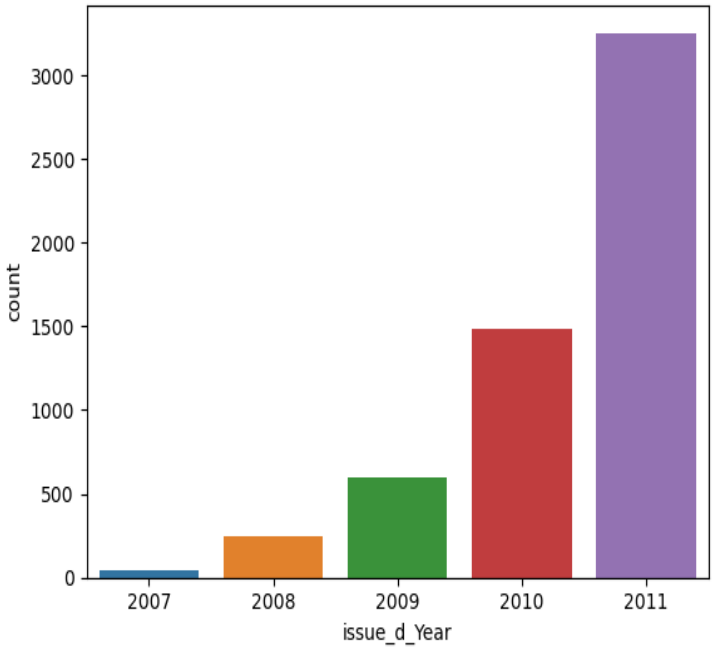
Defaulters % by state



The state CA has the highest number of defaulters among all states.

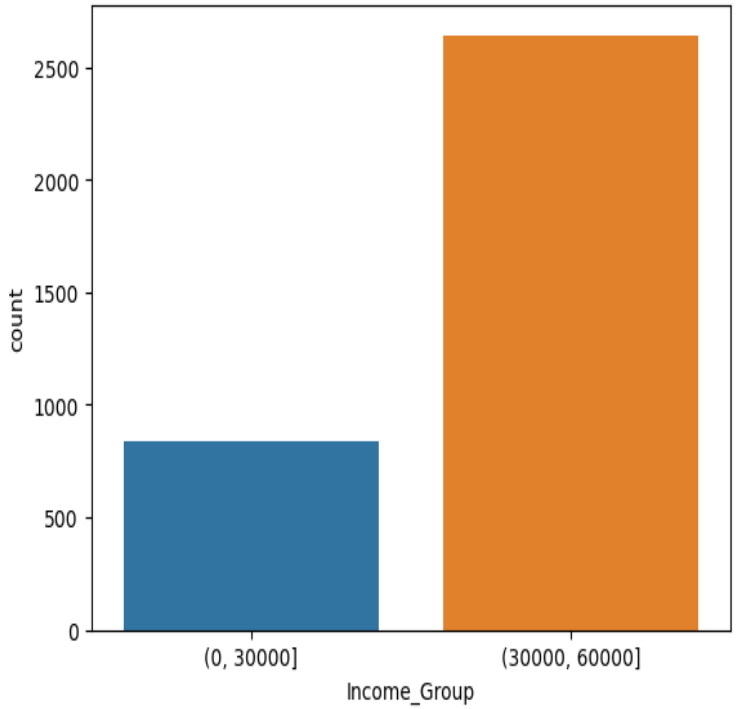
Exploratory Analysis – Key Findings

Defaulters Trend By Year



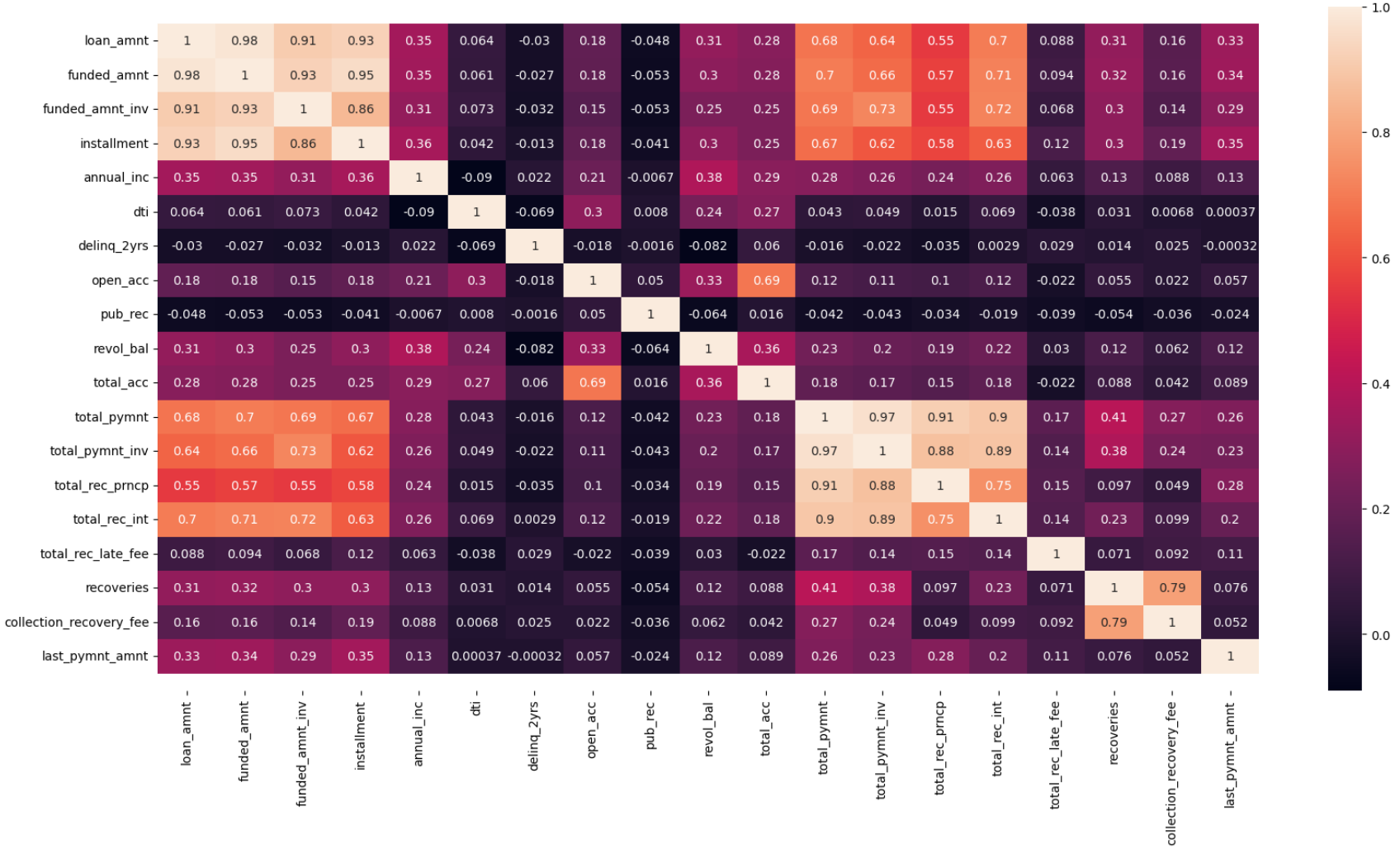
The number of defaulters has been steadily increasing from year to year.

Defaulters by Income Group



Individuals earning more than 30,000 are more inclined to default on their loans compared to those earning less than 30,000

Exploratory Analysis – Key Findings



loan_amnt, funded_amnt, funded_amnt_inv& installment columns are highly correalted

2. total_pymmnt, total_payment_inv,total_rec_prncp,total_rec_int are moderatly correlated to point 1 columns

3. for the further analysis we could reduce the features which are highly correlated

Conclusion

- Lending Club has doubled its expansion in issuing loans every year
- The last two quarters of the year are busier compared to the first two quarters.
- Around 85% of the loans issued are Fully paid 15% of loans issued are defaulted
- Lending Club issue loans only on 2 terms. 36 months and 60 months
- From the above analysis and charts, The Loan amount range from 500 - 35000 , most no.of loans issued are around the amount 10000
- Loans with a 36-month term have a higher default rate compared to those with a 60-month term.
- Loans with grades B, C, and D have a higher default rate compared to loans with other grades.
- The top 5 subgrades with the highest default rates are B5, B3, C1, B4, and C2.
- Borrowers who own their homes have a lower default rate compared to those who do not own a home.
- The state CA has the highest number of defaulters among all states.
- Individuals earning more than 30,000 are more inclined to default on their loans compared to those earning less than 30,000

Thank
You